

EÖTVÖS LORÁND TUDOMÁNY EGYETEM
TERMÉSZETTUDOMÁNYI KAR

**Spatially Embedded Real-World Networks:
Structure and Dynamics
from Global to Urban Scales**

KALLUS ZSÓFIA

Doktori értekezés tézisei

Témavezető: Vattay Gábor, PhD
Egyetemi tanár, az MTA doktora
Komplex Rendszerek Fizikája Tanszék
tanszékvezetője

Fizika Doktori Iskola
Iskolavezető: Tél Tamás, PhD
Egyetemi tanár, az MTA doktora

Statisztikus Fizika, Biológiai Fizika és
Kvantumrendszerek Fizikája Doktori Program
Programvezető: Kürti Jenő, PhD
Egyetemi tanár, az MTA Doktora



Budapest, 2018

Bevezető

Egy *komplex rendszer* természeténél fogva kihívást jelent az egyszerűsítő, modellező látásmódnak. Annak ellenére, hogy sok hasonló építőelemből tevődik össze, melyeket egyszerű kölcsönhatások vezérelnek, a rendszer mint egész váratlan tulajdonságokat képes felmutatni. Az absztrakt gráfok sokoldalú eszközét adják az ilyen összekötött fizikai rendszerek modellezésének – ahogyan azt a *hálózattudomány* iránti fokozott érdeklődés is mutatja. Szerkezeti és dinamikai tulajdonságaik egyaránt érdekesek, és gyakran nem-triviális folyamatokra az alap rendszer statikus tulajdonságai nyújtanak magyarázatot. A *valós világ komplex rendszerei* addicionális kihívásokat hordoznak magukban. Empirikus elemzésük a speciálisan megtervezett mérések és modellezés kombinációját teszi szükségessé, így kutatásuk gyakran *interdiszciplináris* csoportok végzik.

A *térbeli hálózatok* vagy a földrajzi térbe, vagy egy olyan absztrakt geometriai térbe ágyazottak, mely rendelkezik megfelelő távolság definícióval. A valós példák rendkívüli sokszínűséget mutatnak. Közülük sok kötődik nagyvárosokhoz – mint például az elektromos hálózatok, melyek az áramellátásért felelősek, az országokat összekötő kereskedelmi hálózat, a városokat összekötő autópálya hálózata, a repterek hálózata, melyet a világot keresztülszelő repülőjáratok kötnek össze, vagy akár az emberi érintkezések hálózata, melyet járványok kórokozói használhatnak világ körüli utazáshoz.

Ezen nagyméretű hálózatok *empirikus kutatása* sokszor korlátozódik kisebb modell rendszerek vagy részrendszerek tanulmányozásán alapuló következtetésekre. Nagy-skálájú hálózatok esetén pedig sokszor van szükség nem nyilvános adatok elemzésére, ami nem mindig hozzáférhető a kutatók számára. Míg ezen infrastrukturális hálózatok az embereket szolgálják ki, addig egy másik fontos hálózat típus a kapcsolati és kommunikációs hálózatoké, melyeket maguk az egyének alkotnak. Ezek esetében a térbeli elhelyezkedés szintén csak kisebb vagy kísérleti rendszerek esetén volt elérhető – mért adatok hiányában eloszlás modelleket használhatunk, ilyenek a közelségen alapuló modellek, mint például a népszerű gravitációs modellek családja is.

Az elmúlt években empirikus kutatásokra interdiszciplináris csoportok jöttek létre azal a céllal, hogy feltérképezzék alapvető, önszerveződő rendszerek hálózati modelljét, mint például az internet vagy a hálózatosodott társadalmak nagyvárosai. Ezen rendszerek gyors növekedését a társadalom számára való nagyfokú hasznosságuk segítette elő. Így mostanra nélkülözhetetlenné váltak, és szükséges jellemzőik feltárása és például a károkkal szembeni hatékony védelmük elősegítése. Ilyen kutatások során először a topológiát feltáró méréseket kell megtervezni, melyek segítségével elkészíthető a rendszer egy gráf absztrakciója, meghatározható az összeköttetések és a csomópontok relatív súlyozása. A modell kiegészíthető geolokációs mérésekkel is, hogy a rendszer teljes vagy részleges térbeli begyázását adjuk.

Ezen kutatási projektek új mérési rendszereket és szolgáltatásokat hoztak létre, hogy mások is egyfajta virtuális lencsén keresztül nézhessék a kérdéses rendszert – épp mint ahogyan egy asztrofizikus használna egy távcsövet, hogy információt gyűjtsön a körülöttünk lévő Univerzumból. Az asztrofizikai kutatások nagy mennyiségű mérési adatainak könnyű gyűjtése, kezelése, feldolgozása és kereshetősége érdekében tudományos adat központok jöttek létre. Ezek az ún. Virtuális Obszervatóriumok rutinszerűen használnak statisztikus elemzésekhez térbeli tulajdonságokat, és így mintájukra az újonnan

nagy mennyiségű adattal találkozó kutatók, mint a hálózattudomány művelői, új típusú virtuális obszervatóriumokat hoztak létre. Példák erre az internet hálózat mérései, a várostudomány vagy az online kapcsolati hálózatok kutatása. Utóbbi esetében a személyes eszközökön alapuló kollektív érzékelés fejlődését egy csendes forradalom követte a személyes vonatkozású pozíció adatok társadalmi megítélésében. Ezek együttesen tették lehetővé, hogy kutatók minden eddiginél nagyobb mennyiségű nyíltforrású adatot elemezzenek.

Doktori tanulmányaim során nagy kiterjedésű, földrajzilag beágyazott, valós hálózatok vizsgálatát tűztem ki célul. A tézis három része három ilyen fizikai rendszerhez kapcsolódik: a router szintű internet hálózathoz, kibertér globális közösségéhez, melyek tagjait a Twitter online platform köti össze, ill. Nagy-London régió városlakóihoz, akik aktívan használják egy mobilkommunikációs hálózat tornyait. Független tudományos projekteken vehettem részt, a korábban leírt felderítő méréseket végző csapatokkal kollaborációban. Megállapításainkat öt fejezetben mutatom be. Önálló eredményeim a hálózati mérések empirikus elemzéséből nyert tudással gazdagítják ismereteinket a vizsgált rendszerekről.

I - Internet hálózati mérések

1. Likelihood leképezés probabilisztikus geolokációhoz

Az első fejezet összefoglalót ad korábbi interneten végzett topológia felderítő kísérlet sorozatokról, ill. a router szintű hálózat csomópontjainak térbeli beágyazásáról. Az extenzív hálózati tomográfias mérések a PlanetLab végponti gépeiről végezték. A felmérés az inter-PlanetLab minta hálózat térbeli beágyazásával folytatódott, ami egy statisztikus modell alapú, aktív geolokációs módszerrel történt. Ugyanezt a valószínűségi geolokációs módszert más, például tematikus cím felderítő mérősorozatok esetén is alkalmazhatjuk. Ezen kísérletek mindegyike több ezer átfedő térbeli eloszlást eredményezhet.

A fejezet eredményei:

- T1. Bevezettem egy térbeli aggregálási módszert, ami a térbeli bizonytalanság statisztikus mértékeinek egy családján alapszik. Hatékonyan alkalmaztam átfedő térbeli eloszlások aggregálására. A térbeli likelihood térképeket különböző, nagy volumenű kísérletsorozatok eredményének kiértékelésénél használtam fel. Megmutattam, hogy a módszer alkalmas globális lefedettséget adó nagyvolumenű célpont halmazok város szintű csoportosulási struktúráját felfedni. [1]

II - Egy globális közösségi hálózat

2. Geopolitikai régiók effektív magja és perifériája

A második fejezet áttekintést adott a Twitter közösségi platform publikus fiókjaiból álló kapcsolati hálózat korábban végzett felderítéséről. Lehetőségünk van az ismert helyről posztoló egyéni felhasználók alapján a világ országai és államai közötti kapcsolati hálót előállítani az adminisztrációs határok szerinti aggregálással. Az egymást említő és követő felhasználói kapcsolatokat aggregálva súlyozott gráf reprezentációkat kapunk, magas szintű átlagoltsággal.

A fejezet eredményei:

- T2. Elvégeztem a globális regionális kapcsolati hálózat szerkezeti elemzését két súlyozott, irányított regionális gráf reprezentáció alapján. Vírusterjedési szimuláció segítségével meghatároztam az egyes csomópontok effektív összekötöttségét és globális fertőzési erejét. K-shell dekompozícióval a hálózat magjától mért absztrakt távolságok meg lettek határozva. Megmutattam, hogy kombinálva a két típusú csúcs centralitás mértéket, a globális régiós hálózatnak az effektív mag-periféri hálózata sikeresen feltárható. [2]

3. Egy szociális hálózaton terjedő „világjárvány” rejtett mintázata

A harmadik fejezet összefoglalót ad arról, hogyan fedték fel historikus világjárványok egy kiterjedt halmazának rejtett geometriáját. Ehhez a szerzők a világ reptereit összekötő hálózat földrajzi linkhosszait helyettesítették egy hálózati effektív távolsággal¹. Hasonlóan vizsgálva az információ utazását egy közösségi hálózaton, a regionális kapcsolati gráfokat effektív távolságon alapuló beágyazáshoz használtam. Twitter üzeneteket és a Google keresőmotor feljegyzéseit használtam a térben és időben leírt hír trajektória előállításához. A kategóriájának elsőjeként rekordot felállító ?Gangnam Style? videó, egy kisebb közönségtől indulva ért el világméretű lefedettséget – ez pedig a terjedési folyamatot kivételesen alkalmassá teszi a fertőzések világjárványaival való összehasonlításra.

A fejezet eredményei:

- T3. Adaptáltam az effektív távolság módszerét kommunikációs hálózaton való hír terjedés modellezésre. Bemutattam a regionális kapcsolati hálózat egy magas valószínűségű hírforráspont körüli absztrakt beágyazását. Megmutattam, hogy ez a folyamat a gráfon, a hálózatok által hajtott vírusterjedési jelenségekhez hasonlóságot mutat. A rendszer internál eredményeit külső méréssel meghatározott érkezési idő becslésekkel vettem össze ugyanazon beágyazáson, és megmutattam, hogy a statisztikus viselkedés változatlanul hasonló marad. [3]

¹DOI: 10.1126/science.1245200

4. A kibertér közösségeinek térbeli elemzése

A negyedik fejezet az egyének ismeretségi gráfjának egy alacsonyabb szintű földrajzi aggregálását mutatja be, melyben az elegendően kicsi és egyenlő területű HEALPix felosztás celláit használtam. Így előállítható nagy kiterjedésű, térbe ágyazott gráf reprezentációk egy sorozata – az alapul szolgáló online hálózat összeköttetéseinek finom felbontású részleteit megőrizve. Ez a korábban vizsgált hálózatoknál sokkal nagyobb kiterjedésű, és lefedettségű adathalmazt eredményezett.

A fejezet eredményei:

- T4. Modularitás alapú skaság szimulációt alkalmaztam egy nagy kiterjedésű, súlyozott gráfokból álló sorozat klaszterezési elemzéséhez, és megmutattam, hogy moduláris topológiai struktúrájuk van.
- T5. Bemutattam az azonosított közösségek földrajzi térbe vetített projekcióit, és elemeztem ezek térbeli kompaktságát és távolba nyúló befolyását. Megmutattam, hogy általánosan erős kompaktság jellemzi őket, és emellett megtaláltam a nyelvi korlátok és politikai konfliktus zónák szétartó hatásának lenyomatait is.
- T6. Bemutattam két particionálási típus összehasonlító elemzését: az adminisztrációs és a hálózat-alapú felosztásokét, és megmutattam, hogy a szociális kapcsolatok feltárt klaszterei magasabb minőségű kohéziót ad, mint a geopolitikai határvonalakon alapuló felosztás. [4]

III - Városi mobilitás

5. Előrejelezhetőség és városi lokális aktivitás mintázatok

Az ötödik fejezet összefoglalást ad arról, hogyan változtatta át az erősen hálózatosodott városi lakosság a kommunikációs tornyokat az emberi aktivitás lokális szintjének mérésére alkalmas szenzorhálózattá. A helyi aktivitás idősorok néhány meghatározó heti mintázat szerinti klaszterbe particionálhatóak – például üzleti vs. lakossági körzetek. (Ezekre a klaszterekre gondolhatunk a városi mobilitás funkcionális 'klíma' zónáiként.)

A fejezet eredményei:

- T7. Bemutattam városi helyszínek egy előrejelezhetőségi elemzését, a lokális aktivitási idősorok szabályos és szabálytalan részekre való bontása alapján. Elemeztem az előrejelezhetőség határait Nagy-London régió erősen szabályos emberi aktivitás mintázataira. (Az előző analógiát kiterjesztve, a városnak ezen nézete a helyi 'időjárás' idősorok változékonyságának felel meg.) Prediktabilitási mérőszámok segítségével megmutattam a térbeli aggregálás különböző szintjeinek simító hatását és a funkcionális klaszterek hatékonyságát. Megmutattam, hogy a helyszínek aktivitási volumenjeit a helyi különleges események zajszinttől való megkülönböztetésére lehet használni. használhatóak különböző méréstípusoknl egyaránt.

Egyúttal azonosítottunk különleges eseményeket, melyek a tipikus zajszintektől elváló mintázatot mutatnak, és különböző aktivitás típusok idősoraiban hagytak lenyomatot. [5]

Publikációk

A dolgozat alapjául szolgáló publikációk:

- [1] Zs. Kallus, P. Hágá, P. Mátray, G. Vattay, S. Laki. *Complex Geography of the Internet Network*. ACTA PHYSICA POLONICA B 42:(5) pages 1057-1069. 2010.
- [2] Zs. Kallus, N. Barankai, D. Kondor, L. Dobos, T. Hanyecz, J. Szüle, J. Stéger, T. Sebők, G. Vattay, I. Csabai. *Regional properties of global communication as reflected in aggregated Twitter data*. In: CogInfoCom 2013, pages 429-434. 2013.
- [3] Zs. Kallus, D. Kondor, J. Stéger, I. Csabai, E. Bokányi, G. Vattay. *Video Pandemics*. In: Ict Innovations 2017: Data-driven Innovation, pages 3-12. Springer. 2017.
- [4] Zs. Kallus, N. Barankai, J. Szüle, G. Vattay. *Spatial Fingerprints of Community Structure in Human Interaction Network for an Extensive Set of Large-Scale Regions*. PLOS ONE 10:(5) Paper e0126713. 2015.
- [5] D. Kondor, S. Grauwin, Zs. Kallus, I. Gódor, S. Sobolevsky, C. Ratti. *Prediction limits of mobile phone activity modeling*. *Prediction limits of mobile phone activity modelling*. Royal Society open science, 4(2), 160900. 2017.

Egyéb publikációk:

- 6. A. Bodor, L. Diósi, Zs. Kallus, T. Konrad. *Structural features of non-Markovian open quantum systems using quantum chains*. PHYSICAL REVIEW A 87:(5) Paper 052113 7 p. 2013.
- 7. L. Dobos, J. Szüle, T. Bodnár, T. Hanyecz, T. Sebők, D. Kondor, Zs. Kallus, J. Stéger, I. Csabai, G. Vattay. *A multi-terabyte relational database for geo-tagged social network data*. In: CogInfoCom 2013. pp. 289-294. 2013.
- 8. D. Kondor, I. Csabai, L. Dobos, J. Szüle, N. Barankai, T. Hanyecz, T. Sebők, Zs. Kallus, G. Vattay. *Using Robust PCA to estimate regional characteristics of language use from geo-tagged Twitter messages*. In: CogInfoCom 2013. pp. 393-398. 2013.
- 9. I. Gódor, Zs. Kallus, D. Kondor, S. Grauwin, S. Sobolevsky. *The Signatures of City Life*. Ericsson Mobility Report, November 2014, pp 28-29. 2014.
- 10. T. Sebők, Zs. Kallus, S. Laki, P. Mátray, J. Stéger, J. Szüle, G. Vattay. *Network Measurement Virtual Observatory: An Integrated Database Environment for Internet Research and Experimentation*. In International Conference on Testbeds and Research Infrastructures (pp. 65-74). Springer. 2014.
- 11. Zs. Kallus, D. Kondor, I. Csabai, G. Vattay. *Uncovering the hidden geometry of world-wide online news dynamics*. NETSCI 2015, (Accepted, not presented), International School and Conference on Network Science, Spain. 2015.
- 12. I. Gódor, P. Jonsson, Zs. Kallus, D. Kondor. *The Digital Signatures of Sport*. Ericsson Mobility Report June 2015, pp 20-23. 2015.